

Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med* 2014;370:2408-17. DOI: 10.1056/NEJMoa1401268

SUPPLEMENTARY APPENDIX

Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing

Michael R. Wilson, M.D., Samia N. Naccache, Ph.D., Erik Samayoa, B.S., C.L.S., Mark Biagtan, M.D., Hiba Bashir, M.D., Guixia Yu, B.S., Shahriar M. Salamat, M.D., Sneha Somasekar, B.S., Scot Federman, B.A., Steve Miller, M.D., Ph.D., Robert Sokolic, M.D., Elizabeth Garabedian, R.N., M.S.L.S., Fabio Candotti, M.D., Rebecca H. Buckley, M.D., Kurt D. Reed, M.D., M.G.I.S., Teresa L. Meyer, R.N., M.S., Christine M. Seroogy, M.D., Renee Galloway, M.P.H., Sheryl L. Henderson, M.D., Ph.D., James E. Gern, M.D., Joseph L. DeRisi, Ph.D., and Charles Y. Chiu, M.D., Ph.D.

TABLE OF CONTENTS

Supplementary Methods	3
Unbiased next-generation sequencing (NGS) of clinical samples	3
PCR confirmation and phylogenetic analysis	4
Accession numbers	4
Author contributions	6
Figure S1. Confirmatory PCR for detection of <i>Leptospira</i> in the case patient's CSF performed by the University of California, San Francisco (UCSF) laboratory	7
Figure S2. Phylogenetic analysis of gene sequences from representative serovars, strains, and species of <i>Leptospira</i>	8
Figure S3. Determination of <i>Leptospira</i> titers in the case patient's cerebrospinal fluid (CSF) by quantitative polymerase chain reaction (PCR) and standard curve analysis	9
Figure S4. Confirmatory real-time PCR assay for detection of <i>Leptospira</i> in the case patient's CSF performed by the US Centers for Disease Control (CDC) laboratory	10
Table S1. Diagnostic testing for potential microbial causes of the patient's meningoencephalitis	11
Table S2. Summary counts of next-generation sequencing (NGS) reads remaining after each step of the SURPI ("sequence-based ultra-rapid pathogen identification") computational pipeline ¹	13
Table S3. Summary counts of viral reads identified using the SURPI pipeline	14
Table S4. Summary counts of bacterial reads identified using the SURPI pipeline	15
Table S5. PCR primers used for detection and sequencing of the <i>Leptospira</i> strain in the case patient's CSF	16
References	18

SUPPLEMENTARY METHODS

Unbiased next-generation sequencing (NGS) of clinical samples

For the unbiased NGS assay, safeguards to minimize contamination included (1) unidirectional laboratory workflow, (2) nucleic acid extraction using a magnetic bead-based instead of column-based method², and (3) use of an appropriate negative control that was processed identically (serum from an unrelated patient). Nucleic acid was extracted from untreated and DNase-treated CSF (for viral capsid enrichment)^{3,4} and DNase-treated serum using the Qiagen EZ1 robot (Qiagen, Carlsbad, CA). Samples were randomly amplified to cDNA, and NGS libraries were constructed using a modified TruSeq protocol (Illumina, San Diego, CA) as previously described^{5,6}. Library size and concentration were determined using the BioAnalyzer High-Sensitivity DNA kit (Agilent, Santa Clara, CA) and Kapa Universal qPCR kit (Kapa Biosystems, Woburn, MA), respectively. Samples were sequenced on an Illumina MiSeq instrument using 150/350 base pair (bp) paired-end sequencing. The 150 bp first read of the mate pair was analyzed for pathogens prior to completion of the NGS run using a cloud-compatible bioinformatics pipeline for pathogen detection and discovery (Naccache, *et al.*, submitted). The pipeline, named SURPI (“sequence-based ultra-rapid pathogen identification”), had previously been used to identify a novel bunyavirus in the Lone Star tick *Amblyomma americanum*⁴. Briefly, the SURPI pipeline first identifies and computationally subtracts human host sequences using the nucleotide aligner SNAP^{7,8}, followed by SNAP identification of viruses, bacteria, fungi, and parasites by comprehensive mapping of all remaining reads to the National Center for Biotechnology Information (NCBI) nucleotide (nt) reference database (NCBI nt). To detect novel viral pathogens, unmatched reads and *de novo* assembled contigs are mapped using the translated nucleotide aligner RAPSearch⁹ to a viral protein database.

PCR confirmation and phylogenetic analysis

For PCR confirmation and full-length gene sequencing, *Leptospira* primers targeting or encompassing specific genes were designed, either directly from the mapped NGS reads (*lipL32*, *lipL41*, *ompA*, *rpoB*, *secY*) or from a previously published clinical assay¹⁰ (*lipL32*) (Table S5). For estimation of pathogen titer, CSF was analyzed using a SYBR green-based qPCR assay targeting the *secY* gene according to a standard curve method as previously described¹¹ (Fig. S1). Phylogenetic trees were constructed in Geneious version 6.1.5 with PHYML using a TN93 substitution model and Shimodaira-Hasegawa-like branch supports^{12,13}.

Accession numbers

The GenBank accession numbers used for the NGS read mapping and phylogenetic analysis are as follows. **Chromosome, complete sequence:** *L. biflexa* serovar Patoc, CP000786; *L. borgpetersenii* serovar Hardjo-bovis str. L550 chromosome 1, NC008508; *L. borgpetersenii* serovar Hardjo-bovis str. L550, chromosome 2, NC008509. **lipL32 gene:** *L. interrogans* serovar Canicola, KC800990; *L. interrogans* serovar Copenhageni, AE016823; *L. interrogans* serovar Lai, CP001221; *L. kirschneri* serovar Galtoni, AY461914; *L. kirschneri* serovar Grippotyphosa, AF121192; *L. kirschneri* serovar Mozdok, AY461917; *L. noguchii* serovar Fortbragg, AF181556; *L. noguchii* serovar Orleans, AY461914; *L. santarosai* serovar Alexi, AY461926; *L. santarosai* serovar Bakeri, AY461928; *L. santarosai* serovar Canalzonae, AY461922; *L. santarosai* serovar Georgia, AY461923; *L. santarosai* serovar Rioja, AY461921; *L. santarosai* serovar Shermani, AY461927; *L. santarosai* serovar Trinidad, AY461924; *L. santarosai* serovar Tropica, AY461925; *L. weili* serovar Coxi, AY461929; *L. weili* serovar Manhao, AY609331; *L. weili* serovar Vughia, AY461930. **lipL41 gene:** *L. borgpetersenii* serovar Tarassovi, AY622684; *L. santarosai* serovar Alexi, *L. santarosai* serovar Bakeri, AY461966; *L. santarosai* serovar Canalzonae, AY461960; AY461964; *L. santarosai* serovar

Georgia, AY461961; *L. santarosai* serovar Roja, AY461959; *L. santarosai* serovar Shermani, AY461965; *L. santarosai* serovar Trinidad, AY461962; *L. santarosai* serovar Tropica, AY461963. **rpoB gene:** *L. alexanderi* serovar Manhao 3, DQ296129; *L. borgpetersenii* serovar Castellonis, EU747310; *L. borgpetersenii* serovar Tarassovi, EU747307; *L. genomospecies 1* serovar Pingchang, DQ296130; *L. interrogans* serovar Canicola, EU747299; *L. interrogans* serovar Hardjo-prajitno, EU747303; *L. interrogans* serovar Icterohaemorrhagiae, DQ296133; *L. interrogans* serovar Wolffi, EU747308; *L. kirschneri* serovar Bafani, EU747315; *L. kirschneri* serovar Cynopteri, DQ296139; *L. kirschneri* serovar Grippotyphosa, EU747301, *L. kmetyi* serovar Malaysia, AB291211; *L. noguchii* serovar Nicaragua, EU349499; *L. noguchii* serovar Panama, DQ296141; *L. noguchii* serovar Caco, EU349498; *L. santarosai* serovar Shermani, DQ296131; *L. weilii* serovar Cellendoni, DQ296132. **secY gene:** *L. borgpetersenii* serovar Tarassovi, EU358057; *L. santarosai* serovar Alexi, EU358047; *L. santarosai* serovar Atchafalaya, EU358062; *L. santarosai* serovar Atlantae, EU358059; *L. santarosai* serovar Babudieri, EU358056; *L. santarosai* serovar Bagua, EU358046; *L. santarosai* serovar Bakeri, EU358058; *L. santarosai* serovar Balboa, EU357958; *L. santarosai* serovar Beye, EU357981; *L. santarosai* serovar Borincana, EU357978; *L. santarosai* serovar Brasiliensis, EU357959; *L. santarosai* serovar Canalzonae, EU358029; *L. santarosai* serovar Cenepa, EU358054; *L. santarosai* serovar Darien, EU358066; *L. santarosai* serovar Gatuni, EU358061; *L. santarosai* serovar Georgia, EU358033; *L. santarosai* serovar Goiano, EU357995; *L. santarosai* serovar Gorgas, EU357991; *L. santarosai* serovar Guaricura, EU357994; *L. santarosai* serovar Kobbe, EU358024; *L. santarosai* serovar Maru, EU357977; *L. santarosai* serovar Naparuca, EU357971; *L. santarosai* serovar Navet, EU358067; *L. santarosai* serovar Princetown, EU358051; *L. santarosai* serovar Rama, EU358063; *L. santarosai* serovar Rio, Eu358042; *L. santarosai* serovar Sanmartini, EU358050; *L. santarosai* serovar Shermani, EU358055; *L. santarosai* serovar Tabaquite, EU357982; *L. santarosai* serovar Trinidad, EU358035; *L. santarosai* serovar

Tropica, EU358016; *L. santarosai* serovar Vargonicas, EU358011; *L. santarosai* serovar Weaveri, EU358005.

Full-length sequences of the *lipL41*, *rpoB*, and *secY* genes and a partial sequence of the *lipL32* gene from the patient's *Leptospira* strain (*L. santarosai* strain UW) have been deposited in GenBank (accession numbers KJ152438 – KJ152442). Clinical non-human metagenome data corresponding to NGS of the patient's CSF and serum, as well as serum from an unrelated patient used as a negative control, have been deposited in the NIH Sequence Read Archive (accession number SRP035384).

Author contributions

M.R.W., J.G., J.L.D., and C.Y.C. conceived of and coordinated the study. S.N.N., E.S., S.S., and C.Y.C. generated the NGS libraries and analyzed the NGS data. S.N.N., S.F., and C.Y.C. developed the SURPI computational pipeline for rapid pathogen identification from NGS data. M.R.W., E.S., G.Y., and R.G. performed the confirmatory PCR and full-length gene sequencing studies. M.B., H.B., S.M.S., K.R., R.S., E.G., F.C., R.H.B., T.L.M., C.M.S., S.L.H., and J.G. cared for the patient and compiled clinical and pathology data. S.M., R.G., J.G., J.L.D., C.Y.C. contributed reagents/materials/analysis tools. M.R.W. and C.Y.C. wrote the first draft of the manuscript. M.R.W., S.N.N., E.S., M.B., H.B., S.M., R.S., R.H.B., C.M.S., R.G., S.L.H., J.G., J.L.D., and C.Y.C. collectively wrote the paper. All authors vouch for the data and the analysis and decided to publish the paper.

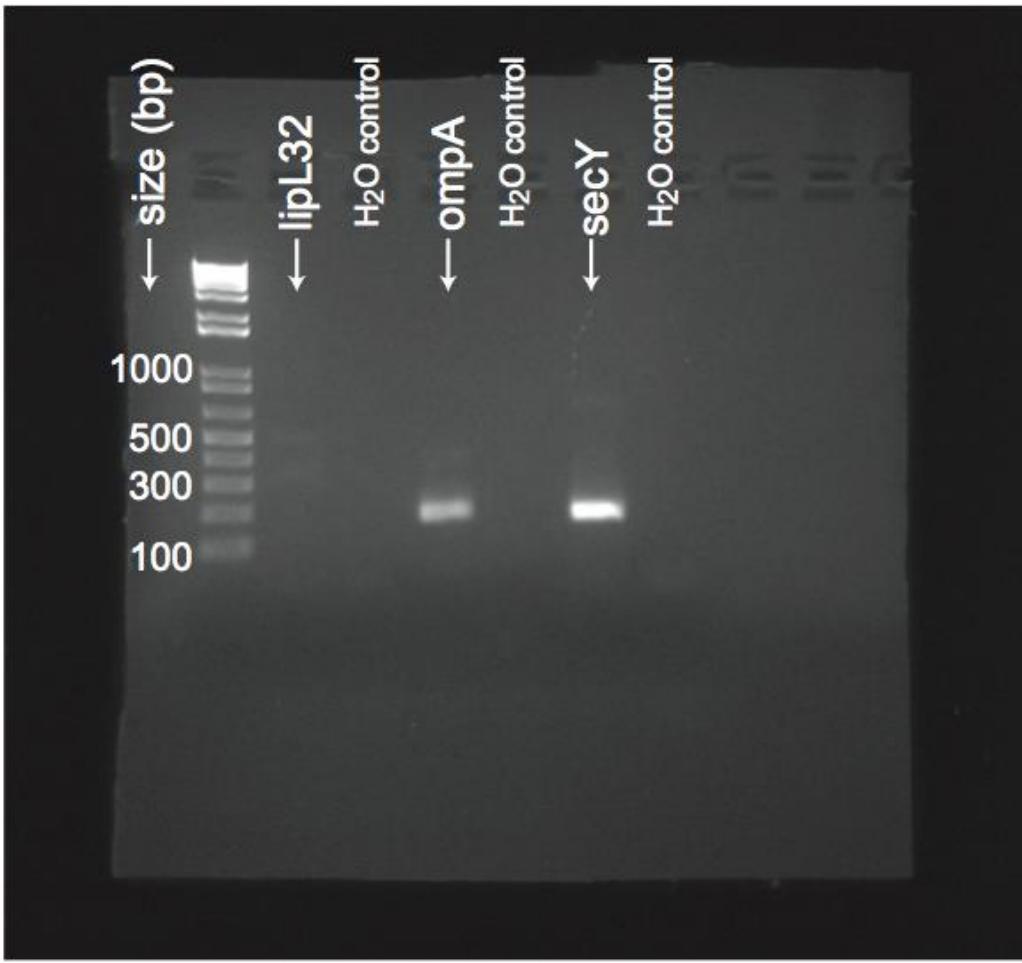


Figure S1. Confirmatory PCR for detection of *Leptospira* in the case patient's CSF performed by University of California, San Francisco (UCSF) laboratory. PCR primers targeting the *ompB* and *secY* genes were designed directly from the sequence reads obtained by next-generation sequencing (NGS). PCR primers targeting the *lipL32* gene were obtained from a previously published CLIA (Clinical Laboratory Improvement Amendments)-validated assay for *Leptospira* detection developed by the Centers for Disease Control (CDC)¹⁰ (also see Fig. S4). Bands of the expected size were confirmed by cloning and Sanger sequencing of the corresponding PCR amplicons. Abbreviations: bp, base pairs; H_2O , water.

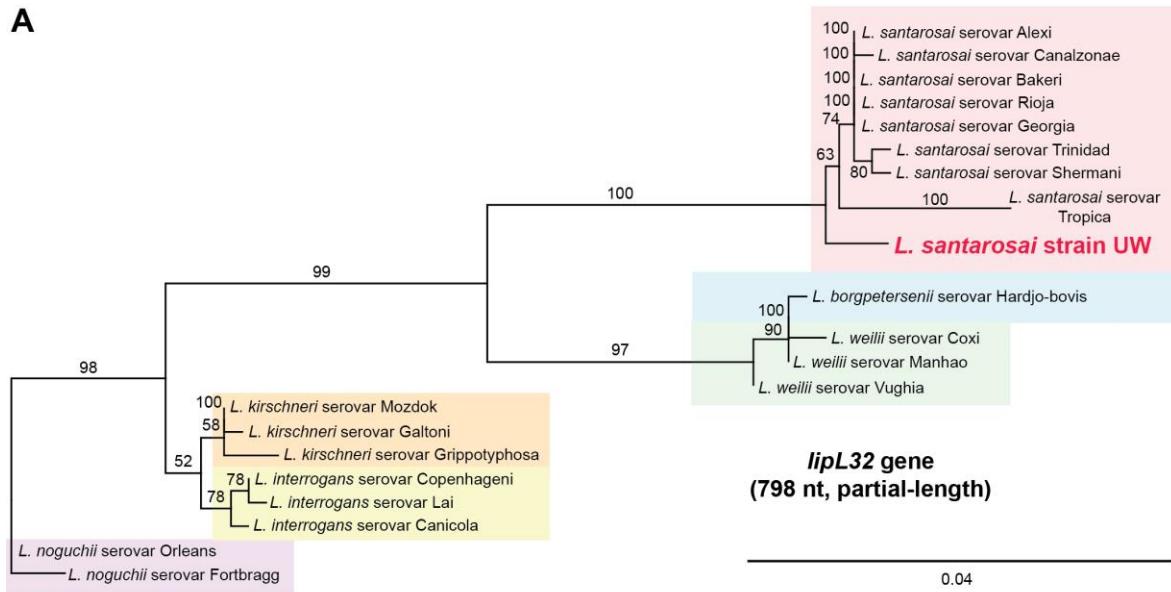
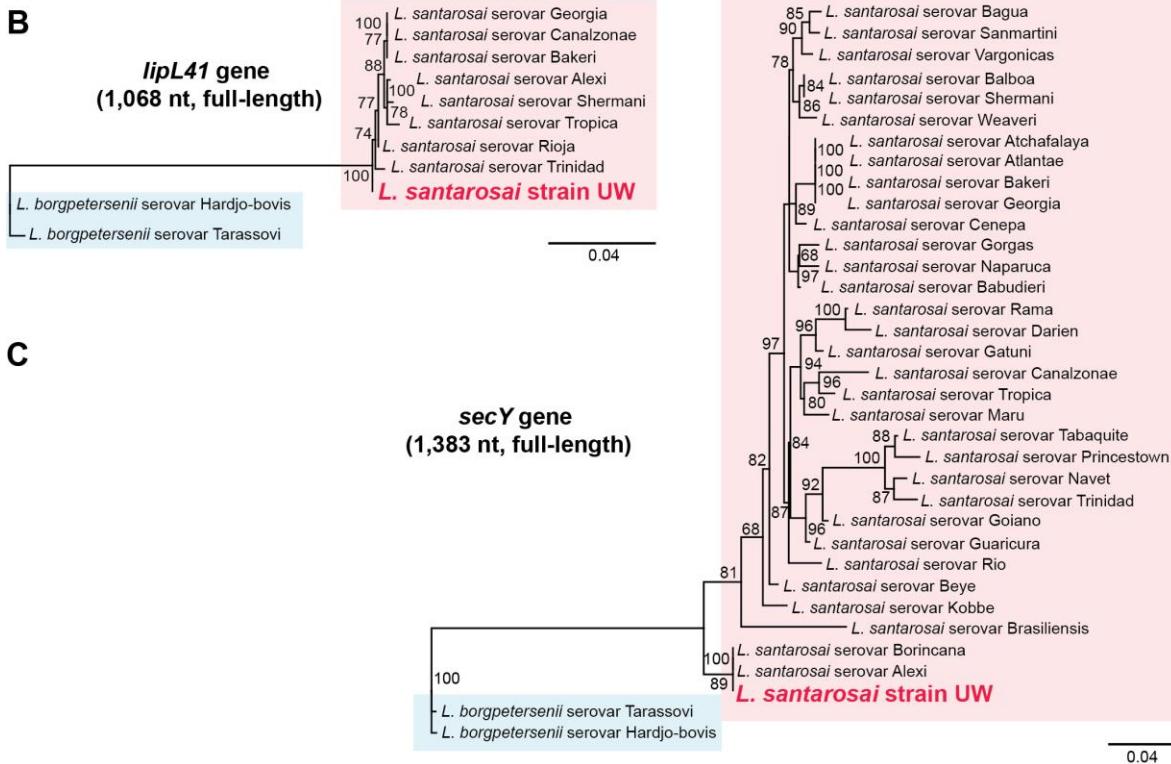
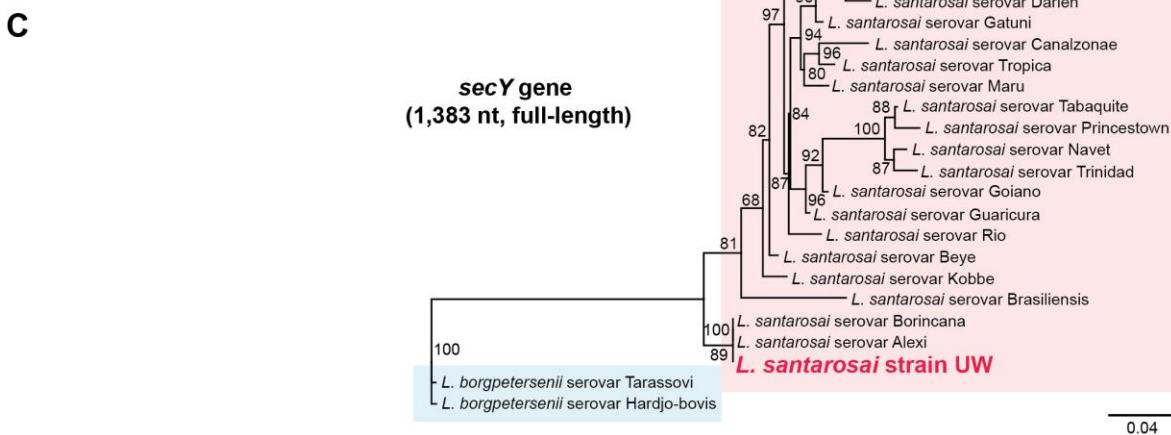
A**B****C**

Figure S2. Phylogenetic analysis of gene sequences from representative serovars, strains, and species of *Leptospira*. (A) *lipL32* gene. (B) *lipL41* gene. (C) *secY* gene. Note that the *lipL32* gene is present exclusively in pathogenic *Leptospira*¹⁰. The strain identified in the case patient with fulminant meningoencephalitis (*L. santarosai* strain UW) is highlighted in boldface red. GenBank accession numbers are provided on page 6. Abbreviations: UW, University of Wisconsin.

***Leptospira* qPCR Assay Targeting secY Gene**

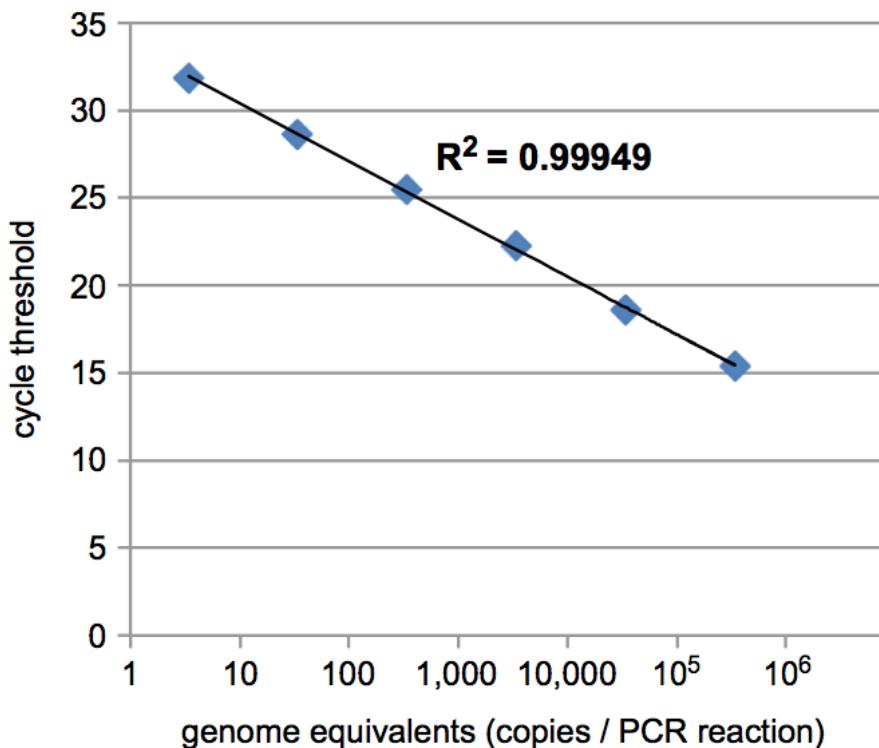


Figure S3. Determination of *Leptospira* titers in the case patient's CSF by quantitative PCR and standard curve analysis. Shown is a log-log plot of the standard curve corresponding to a *Leptospira* SYBR-green quantitative PCR (qPCR) assay targeting the secY gene. The assay was run using 6 serial dilutions of a quantified secY control amplicon in PCR buffer. Each data point is an average of three independent replicates.

CDC Real-Time Assay for Leptospirosis (Stoddard, et al., 2009, Diagn Microbiol Infect Dis 64:247-55)

█ patient extracted DNA (CSF sample aliquot) █ positive *Leptospira* DNA control █ negative “no-template” control

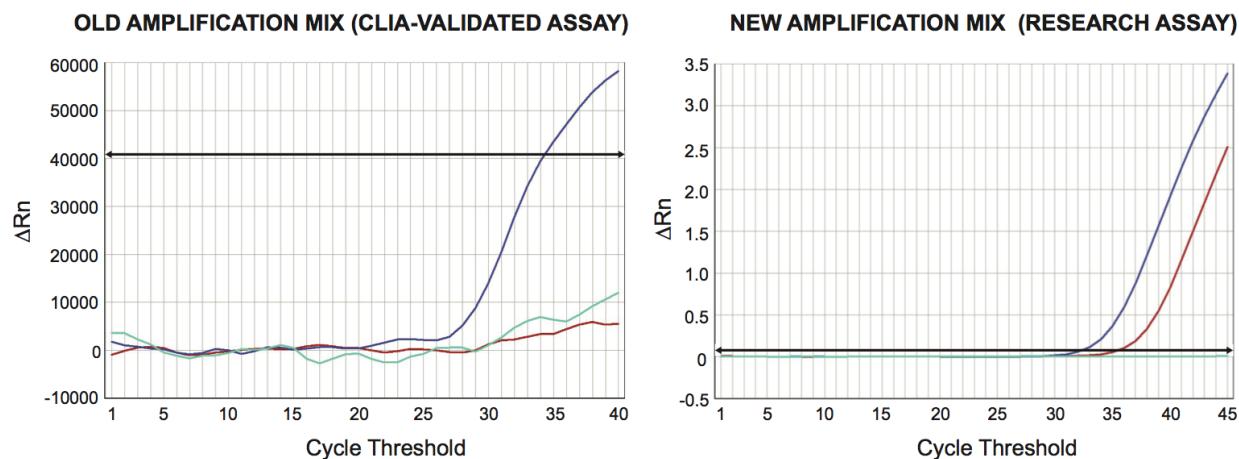


Figure S4. Confirmatory real-time PCR assay for detection of *Leptospira* in the case patient’s cerebrospinal fluid (CSF) performed by the US Centers for Disease Control (CDC) laboratory. PCR amplification curves correspond to extracted CSF DNA from the case patient (red), a positive *Leptospira* DNA control (blue), and a negative “no-template” control (mint green). Note that *Leptospira* amplification from the case patient’s CSF is negative using the Clinical Laboratory Improvement Amendments (CLIA)-validated assay (left) but becomes positive after switching to a new amplification mix (right). The old amplification mix used for CLIA-validated *Leptospira* assay¹⁰ is the Platinum™ PCR SuperMix (Life Technologies), while the new amplification mix is the PerfeCTa™ qPCR ToughMix (Quanta BioSciences). The different results underscore the inadequate sensitivity of the clinically validated *lipL32* PCR assay by the CDC for detection of the patient’s *L. santarosai* infection. The assay had previously been validated using urine and blood and not CSF, and included only 5 *L. santarosai* out of 51 *Leptospira* cultured strains and no *L. santarosai* clinical samples¹⁰, likely explaining why switching to a more robust PCR amplification mix was necessary to boost sensitivity and enable successful detection. The y-axis scales for the fluorescence emission intensity (ΔRn) are different because a ROX reference dye for normalization is absent from the old amplification mix but present in the new amplification mix.

Table S1. Diagnostic testing for potential microbial causes of the patient's meningoencephalitis*.

	Sample Type	First Hosp	Second Hosp	Third Hosp
<i>M. pneumoniae</i> PCR	CSF	—	—	—
<i>Histoplasma / Blastomyces</i> antigen	CSF	—	—	—
<i>Bartonella</i> PCR	CSF	—	—	—
VZV / HHV-6 / HHV8 PCR	CSF	—	—	—
<i>Borrelia burgdorferi</i> PCR	CSF	—	—	—
Adenovirus / CMV / EBV PCR	CSF	—	—	—
HSV-1,2 PCR	CSF	—	—	—
Enterovirus PCR	CSF	—	—	—
Cryptococcal antigen	CSF	—	—	—
Mycobacterial culture	CSF	—	—	—
West Nile Virus IgG/IgM and PCR	CSF	—	—	—
CEV / EEEV / WEEV / SLEV IgG/IgM	CSF	—	—	—
Bacterial / fungal culture	CSF	—	—	—
<i>Toxoplasma gondii</i> PCR	CSF	—	—	—
Powassan virus PCR	CSF	—	—	—
<i>Aspergillus</i> antigen	CSF	—	0.11 (nl <0.5)	—
JCV / BKV / HHV7 PCR	CSF	—	—	—
Viral culture (including mumps culture)	CSF	—	—	—
16S bacterial rRNA PCR	CSF	—	— (x2)	—
Bacterial culture	Brain	—	—	—
Adenovirus / CMV / EBV / VZV PCR	Brain	—	—	—
Enterovirus PCR	Brain	—	—	—
Epstein-Barr virus PCR	Brain	—	—	—
Varicella zoster virus PCR	Brain	—	—	—
EBV / CMV PCR	Plasma	—	—	—
Bacterial culture	Blood	—	—	—
Enterovirus PCR	Serum	—	—	—
Parvovirus B19 / HHV7 / BKV / JCV PCR	Serum	—	—	—
<i>Blastomyces / Histoplasma / Cryptococcus</i> antigen	Serum / Urine	—	—	—
BKV PCR**	Urine	—	—	+
<i>Toxoplasma gondii</i> PCR	Serum	—	—	—
Adenovirus / HTLV-1,II / HIV / HSV-1,2 PCR	Serum	—	—	—
16S bacterial rRNA PCR	Blood	—	—	—
Influenza A, B / RSV PCR	NP swab	—	—	—
Respiratory Viral Panel (Luminex)**	NP swab	—	—	+(rhinovirus)
<i>Mycoplasma pneumoniae</i> PCR	OP swab	—	—	—
Bacterial culture**	Sputum	—	—	+(MRSA)
Enterovirus PCR	Stool	—	—	—

*Microbiological tests were performed by the University of Wisconsin Clinical Laboratory, the Focus Diagnostic Reference Laboratory, and/or ARUP Reference Laboratory. Confirmatory diagnostic testing related specifically to neuroleptospirosis is given in Table 1 of the main text. Abbreviations: hosp, hospitalization; nl, normal; PCR, polymerase chain reaction; ELISA, enzyme-linked immunosorbent assay; IHA, indirect hemagglutination assay; VZV, varicella-zoster virus; HHV, human herpesvirus virus; CMV, cytomegalovirus; EBV, Epstein-Barr virus; HSV, herpes simplex virus; CEV, California encephalitis virus; VEEV, Venezuelan equine encephalitis virus; WEEV, Western equine encephalitis virus; SLEV, St. Louis encephalitis virus; JCV, JC virus; BKV, BK virus; HTLV, human T-cell lymphotropic virus; HIV, human immunodeficiency virus; RSV, respiratory syncytial virus; NP, nasopharyngeal; OP, oropharyngeal; MRSA, methicillin-resistant *Staphylococcus aureus*.

**The findings of BK virus positivity in the urine, rhinovirus positivity in the NP swab, and MRSA positivity in sputum were thought to be incidental and unrelated to the patient's clinical presentation of meningoencephalitis.

Table S2. Summary counts of NGS reads remaining after each step of the SURPI computational pipeline.

number of single-end sequence reads**	case patient			unrelated patient
	Untreated CSF*	DNase-treated CSF*	DNase-treated serum*	DNase-treated serum*
raw reads	3,063,784	2,235,787	2,888,166	2,008,883
after preprocessing [†]	2,924,644	2,224,068	2,881,047	2,006,562
non-human reads	52,621	1,282,953	2,806,178	16,670
nucleotide alignment to NCBI nt database [‡]	17,107	116,532	202,195	6,330
# of reads identified as viral	960	107,016	196,752	107
# of reads identified as bacterial	589	5,233	4,486	3,708
number of paired-end sequence reads**	Untreated CSF*	DNase-treated CSF*	DNase-treated serum*	DNase-treated serum*
	6,127,568	4,471,574	5,776,332	4,017,766
raw reads	5,842,731	4,444,203	5,754,878	4,011,404
after preprocessing [†]	146,608	2,577,989	5,607,069	51,595
non-human reads [‡]	35,126	228,344	399,884	13,359
nucleotide alignment to NCBI nt database [‡]	1,811	208,831	389,038	180
# of reads identified as viral	1,181	10,275	8,929	7,387
# of reads identified as bacterial				

*four NGS libraries were constructed from extracted nucleic acid corresponding to untreated cerebrospinal fluid (CSF) from the case patient (“untreated CSF”), CSF from the case patient pretreated with DNase for virus enrichment^{3,4} (“DNase-treated CSF”), (3) serum from the case patient pretreated with DNase (“DNase-treated serum, case patient”), and (4) serum from a negative control serum sample from an unrelated patient (“DNase-treated serum, unrelated patient”).

**single-end 150 bp (base pair) reads were analyzed after the first 150 sequencing cycles on an Illumina MiSeq™ instrument (~16 hours), while paired-end 150/350 bp reads were analyzed after completion of the run (~40 hours).

[†]the preprocessing step consists of quality filtering, adapter trimming, and removal of low-complexity sequences¹.

[‡]after preprocessing, the SNAP algorithm⁷ is used to remove human reads from the dataset and identify pathogen reads by nucleotide (nt) alignment to human and NCBI nt reference databases, respectively.

Table S3. Summary counts of bacterial reads identified using the SURPI pipeline.

		case patient		unrelated patient	
single-end reads		Untreated CSF*	DNase-treated CSF**	DNase-treated serum**	DNase-treated serum**
Family	Genus	Untreated CSF*	DNase-treated CSF**	DNase-treated serum**	DNase-treated serum**
<i>Leptospiraceae</i>	<i>Leptospira</i>	475	263	0	0
N/A (uncultured bacteria)	N/A	32	870	752	610
<i>Propionibacteriaceae</i>	<i>Propionibacterium</i>	15	296	18	60
<i>Pseudomonadaceae</i>	<i>Pseudomonas</i>	10	589	539	317
<i>Enterobacteriaceae</i>	<i>Escherichia</i>	8	40	220	84
<i>Burkholderiaceae</i>	<i>Burkholderia</i>	7	984	1,296	398
<i>Streptococcaceae</i>	<i>Streptococcus</i>	5	3	0	79
<i>Enterobacteriaceae</i>	<i>Klebsiella</i>	4	0	8	0
<i>Burkholderiaceae</i>	<i>Ralstonia</i>	3	171	387	95
<i>Comamonadaceae</i>	<i>Variovorax</i>	3	117	145	30
<i>Moraxellaceae</i>	<i>Acinetobacter</i>	2	130	68	17
<i>Bacillaceae</i>	<i>Bacillus</i>	2	3	4	35
<i>Acholeplasmataceae</i>	<i>Candidatus Phytoplasma</i>	2	0	0	0
<i>Micrococcaceae</i>	<i>Micrococcus</i>	2	1	0	15
<i>Acidithiobacillaceae</i>	<i>Acidithiobacillus</i>	1	1	37	1
<i>Comamonadaceae</i>	<i>Acidovorax</i>	1	286	61	370
<i>Pasteurellaceae</i>	<i>Aggregatibacter</i>	1	0	0	0
<i>Comamonadaceae</i>	<i>Alicyclophilus</i>	1	4	0	1
<i>Thermoanaerobacteraceae</i>	<i>Caldanaerobacter</i>	1	0	0	0
<i>Comamonadaceae</i>	<i>Comamonas</i>	1	60	15	102
<i>Corynebacteriaceae</i>	<i>Corynebacterium</i>	1	83	49	48
<i>Fusobacteriaceae</i>	<i>Fusobacterium</i>	1	0	0	14
N/A	<i>Niastella</i>	1	0	0	0
<i>Burkholderiaceae</i>	<i>Pandoraea</i>	1	119	53	14
<i>Intrasporangiaceae</i>	<i>Phycicoccus</i>	1	0	1	1
<i>Moraxellaceae</i>	<i>Psychrobacter</i>	1	1	0	0
<i>Alcaligenaceae</i>	<i>Pusillimonas</i>	1	1	0	0
<i>Comamonadaceae</i>	<i>Ramlibacter</i>	1	18	6	4
<i>Rhodobacteraceae</i>	<i>Ruegeria</i>	1	0	0	0
<i>Staphylococcaceae</i>	<i>Staphylococcus</i>	1	2	76	517
<i>Xanthomonadaceae</i>	<i>Stenotrophomonas</i>	1	66	51	14
<i>Oceanospirillaceae</i>	<i>Thalassolituus</i>	1	8	18	20
<i>Xanthomonadaceae</i>	<i>Xanthomonas</i>	1	1	1	2
Other		0	1116	681	860
TOTAL		589	4,117	3,805	2,848

paired-end reads		Untreated CSF*	DNase-treated CSF**	DNase-treated serum**	DNase-treated serum**
Family	Genus	Untreated CSF*	DNase-treated CSF**	DNase-treated serum**	DNase-treated serum**
<i>Leptospiraceae</i>	<i>Leptospira</i>	955	483	0	0
N/A (uncultured bacteria)	N/A	62	1760	1508	1213
<i>Propionibacteriaceae</i>	<i>Propionibacterium</i>	31	543	32	129
<i>Pseudomonadaceae</i>	<i>Pseudomonas</i>	21	1151	1049	652
<i>Enterobacteriaceae</i>	<i>Escherichia</i>	15	83	441	171
<i>Burkholderiaceae</i>	<i>Burkholderia</i>	12	1958	2562	816
<i>Streptococcaceae</i>	<i>Streptococcus</i>	8	7	0	168
<i>Enterobacteriaceae</i>	<i>Klebsiella</i>	6	1	16	0
<i>Comamonadaceae</i>	<i>Variovorax</i>	6	244	252	59
<i>Bacillaceae</i>	<i>Bacillus</i>	5	8	8	70
<i>Acholeplasmataceae</i>	<i>Candidatus</i>	5	0	0	0

	<i>Phytoplasma</i>				
Burkholderiaceae	<i>Ralstonia</i>	4	304	746	167
Moraxellaceae	<i>Acinetobacter</i>	3	277	136	35
Micrococcaceae	<i>Micrococcus</i>	3	2	0	28
Staphylococcaceae	<i>Staphylococcus</i>	3	4	154	995
Xanthomonadaceae	<i>Stenotrophomonas</i>	3	131	97	26
Corynebacteriaceae	<i>Corynebacterium</i>	2	148	93	97
Fusobacteriaceae	<i>Fusobacterium</i>	2	0	0	28
Legionellaceae	<i>Legionella</i>	2	0	22	2
Moraxellaceae	<i>Moraxella</i>	2	82	29	2
Burkholderiaceae	<i>Pandoraea</i>	2	226	108	24
Moraxellaceae	<i>Psychrobacter</i>	2	1	0	0
Alcaligenaceae	<i>Pusillimonas</i>	2	1	0	0
Oceanospirillaceae	<i>Thalassolituus</i>	2	16	39	45
Acidithiobacillaceae	<i>Acidithiobacillus</i>	1	1	71	1
Comamonadaceae	<i>Acidovorax</i>	1	509	152	716
Pasteurellaceae	<i>Aggregatibacter</i>	1	0	0	0
Comamonadaceae	<i>Alicycliphilus</i>	1	10	0	2
Rhodospirillaceae	<i>Azospirillum</i>	1	0	0	3
Thermoanaerobacteraceae	<i>Caldanaerobacter</i>	1	0	0	0
Comamonadaceae	<i>Comamonas</i>	1	128	35	227
Burkholderiaceae	<i>Cupriavidus</i>	1	136	15	53
Comamonadaceae	<i>Delftia</i>	1	255	264	431
Enterobacteriaceae	<i>Enterobacter</i>	1	0	38	29
Enterobacteriaceae	<i>Erwinia</i>	1	0	0	9
Comamonadaceae	<i>Hydrogenophaga</i>	1	12	18	20
Leptotrichiaceae	<i>Leptotrichia</i>	1	0	0	0
Listeriaceae	<i>Listeria</i>	1	0	0	0
Methylobacteriaceae	<i>Methylobacterium</i>	1	0	0	9
N/A	<i>Niastella</i>	1	0	0	0
Intrasporangiaceae	<i>Phycicoccus</i>	1	0	1	2
Flavobacteriaceae	<i>Psychroflexus</i>	1	0	0	5
Comamonadaceae	<i>Ramlibacter</i>	1	38	12	7
Rhizobiaceae	<i>Rhizobium</i>	1	33	193	35
Rhodobacteraceae	<i>Ruegeria</i>	1	0	0	0
Enterobacteriaceae	<i>Shigella</i>	1	0	1	3
Xanthomonadaceae	<i>Xanthomonas</i>	1	1	2	3
Other		0	1722 [†]	835 [‡]	1105 [¶]
TOTAL		1,181	8,553	8,094	6,282

*ranked by number of bacterial reads assigned to a given family / genus in the untreated CSF from the case patient (“untreated CSF”)

**pretreatment with DNase (“DNase-treated”) enriches for highly conserved bacterial 16S ribosomal RNA, which hinders accurate species identification of bacteria

[†]representing 57 additional bacterial genera out of 68

[‡]representing 32 additional bacterial genera out of 41

[¶]representing 99 additional bacterial genera out of 118

Table S4. Summary counts of viral reads identified using the SURPI pipeline.

		case patient			unrelated patient
single-end reads					
Family	Genus	Untreated CSF	DNase-treated CSF*	DNase-treated serum*	DNase-treated serum*
Anelloviridae	N/A	594	60,238	123,668	49
Anelloviridae	<i>Alphatorquevirus</i>	269	34,629	43,878	22
Anelloviridae	<i>Betatorquevirus</i>	96	12,121	29,111	3
Inoviridae*	<i>Inovirus</i>	1	24	52	9
Podoviridae*	<i>T7likevirus</i>	0	4	43	23
Herpesviridae	<i>Roseolovirus</i>	0	0	0	1
TOTAL		960	107,016	196,752	107
paired-end reads					
Family	Genus	Untreated CSF	DNase-treated CSF*	DNase-treated serum*	DNase-treated serum*
Anelloviridae	N/A	1,114	118,521	243,727	80
Anelloviridae	<i>Alphatorquevirus</i>	512	66,693	87,906	30
Anelloviridae	<i>Betatorquevirus</i>	182	23,554	57,205	6
Inoviridae**	<i>Inovirus</i>	3	57	113	18
Podoviridae**	<i>T7likevirus</i>	0	6	87	45
Herpesviridae	<i>Roseolovirus</i>	0	0	0	1
TOTAL		1,811	208,831	389,038	180

*pretreatment with DNase (“DNase-treated”) is used to enrich NGS libraries for reads from encapsidated viral genomes^{3,4}.

**bacteriophages

Table S5. PCR primers used for detection and sequencing of case patient's *Leptospira* strain.

GENE REGION	PRIMER NAME	PRIMER SEQUENCE	SIZE
PCR confirmation			
<i>ompA</i>	lepto-ompA-102F	AAGCTTCAATTCCGGGCG	160bp
	lepto-ompA-293R	CCCCTGCAAACCCGTAGAT	
<i>secY</i> (used for quantification of <i>Leptospira</i> titer)			
	lepto-secY-39F	ATCTTCGCGCTTGGGATCAT	190bp
	lepto-secY-228R	TTGCGAGTTGGATTACCGCT	
<i>lipL32</i> (CLIA-validated PCR) ⁷			
	lepto-lipL32-F	AAGCATTACCGCTTGTGGTG	242 bp
	lepto-lipL32-R	GAACCTCCATTTCAGCGATT	
<i>lipL32</i> (outside of region encompassed by the CLIA-validated PCR)			
	lepto-lipL32-outside-F	CTCCGTTGCCTCTTGCAAGC	330bp
	lepto-lipL32-outside-R	GAACGCATCACTTACTAAATCTCCG	
<i>Leptospira</i> gene sequencing			
<i>rpoB</i>	lepto-rpoB-F	CGG ACA AGG GAA GAC TTC GT	~8 kb
	lepto-rpoB-R	TCG CGT CTT TCG GTC TTC TC	
<i>lipL32</i>	lepto-lipL32-F	CCGAACATCAAACCTGATCCC	~3 kb
	lepto-lipL32R	GTTTGAACGGTAGCGGGAATTCC	
<i>lipL41</i>	lepto-lipL41-F	TGGTGTTACTCTCCATGAGAAAA	~2 kb
	lepto-lipL41-R	TCAGGGTCGTGGATAAGAGTACT	
<i>secY</i>	lepto-secY-F	CAAGAGTGCTTTGAAACCGAGT	~4 kb
	lepto-secY-R	TCTTAATGTTCAACCCTCTGGCA	
	secY-gap-1F	TCCCGGTCCAGATACCAATTAC	close gaps
	secY-gap1R	ACCGCTTCCTCAACAATACG	close gaps
	secY-gap2F	TCTGAGGGATAGAAAGTCTTGCAC	close gaps
	secY-gap2R	TCCGTTCTCGCAGATCTGGTATC	close gaps
	secY-gap3F	TACCGCTAAGGATTGAATCGCAC	close gaps
	secY-gap3R	TCCGTATGGGTACGCACATTAC	close gaps
	secY-gap4F	TCTATGTTCCGGTCAAACCTG	close gaps
	secY-gap4R	AGGGATGTTGGATTCAAGAC	close gaps

REFERENCES

1. Naccache SN, Federman S, Veeraraghavan N, et al. A cloud-compatible bioinformatics pipeline for ultra-rapid pathogen identification from next-generation sequencing of clinical samples. (submitted and under review) 2014.
2. Naccache SN, Greninger AL, Lee D, et al. The perils of pathogen discovery: origin of a novel parvovirus-like hybrid genome traced to nucleic Acid extraction spin columns. *Journal of virology* 2013;87:11966-77.
3. Allander T, Emerson SU, Engle RE, Purcell RH, Bukh J. A virus discovery method incorporating DNase treatment and its application to the identification of two bovine parvovirus species. *Proc Natl Acad Sci U S A* 2001;98:11609-14.
4. Swei A, Russell BJ, Naccache SN, et al. The genome sequence of Lone Star virus, a highly divergent bunyavirus found in the Amblyomma americanum tick. *PloS one* 2013;8:e62083.
5. Greninger AL, Chen EC, Sittler T, et al. A metagenomic analysis of pandemic influenza A (2009 H1N1) infection in patients from North America. *PloS one* 2010;5:e13381.
6. Grard G, Fair JN, Lee D, et al. A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLoS pathogens* 2012;8:e1002924.
7. Zaharia M, Bolosky WJ, Curtis K, et al. Faster and more accurate sequence alignment with SNAP. *arXiv* 2011;1111.5572.
8. Zaharia M, Bolosky B, Curtis K, et al. Alignment in a SNAP: Cancer Diagnosis in the Genomic Age. *Lab Invest* 2012;92:458A-A.
9. Zhao Y, Tang H, Ye Y. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* 2012;28:125-6.
10. Stoddard RA, Gee JE, Wilkins PP, McCaustland K, Hoffmaster AR. Detection of pathogenic Leptospira spp. through TaqMan polymerase chain reaction targeting the LipL32 gene. *Diagn Microbiol Infect Dis* 2009;64:247-55.
11. Yu G, Greninger AL, Isa P, et al. Discovery of a novel polyomavirus in acute diarrheal samples from children. *PloS one* 2012;7:e49449.
12. Guindon S, Delsuc F, Dufayard JF, Gascuel O. Estimating maximum likelihood phylogenies with PhyML. *Methods Mol Biol* 2009;537:113-37.
13. Kearse M, Moir R, Wilson A, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 2012;28:1647-9.